Anticipatory Detection of Depression from Social Media Platforms Using Powerful Deep and Machine Learning Algorithms

1st Hardik Ghuge-232IT013
Information Technology
National Institute of Technology
Karnataka Surathkal, India 575025
ghugehardikbaban,232it013@nitk.edu.in

2nd Dhananjay Kumar-232IT012
Information Technology
National Institute of Technology
Karnataka Surathkal, India 575025
dhananjaykumar.232it012@nitk.edu.in

Guide Anand Kumar M
Information Technology
National Institute of Technology
Karnataka Surathkal, India 575025
anand.191it104@nitk.edu.in

***Abstract*—In contemporary times, both children and adults are grappling with mental health issues such as depression and anxiety, significantly impacting their lives. However, awareness about mental health problems remains inadequate in certain regions. Depression stands out as a prominent factor contributing to suicides globally, with a substantial number of cases going undiagnosed and untreated. The timely gathering of emotional responses has gained paramount importance, particularly with the widespread use of social media platforms like Instagram, Twitter (referred to as X), Facebook, Snapchat, etc., and the increasing internet user base.**

The precarious nature of mental illnesses has spurred ongoing research in this area. With the evolution of machine learning and deep learning, there is an opportunity to leverage relevant sample data for the early detection of depression. The Child Mind Institute, focusing on teenagers and their day-to-day interactions with mobile phones and social media, has observed a concerning impact on mental well-being and cognitive development at an early age. Their research indicates that 71% of kids experience the effects of anxiety and depression.

The CEO of Social Awakening advocates for a mindful use of technology on social media, urging parents to delay providing smartphones to children until at least the 8th grade. Recognizing depression symptoms at an early stage is crucial, as failure to do so may lead a person to a state of distress from which they may struggle to recover. Social networking platforms serve as conduits for sharing various data and routines with others, offering valuable insights into the mindset of a person experiencing depression.

Drawing from past successes, machine learning algorithms have proven to be effective in predicting the prevalence of depression based on early symptoms and social media activity. The objective of this research is twofold: the first part centers around the temporal and writing patterns of content, while the second part delves into linguistic clues by analyzing the text or tweets shared. The proposed approach highlights the feasibility of using RNN and LSTM to achieve early recognition of depression across various social media platforms, including platforms like Facebook comments.

***Index Terms*—Depression, Machine Learning, Deep Learning,Accuracy.**

I. INTRODUCTION

Depression is a prevalent and potentially lethal medical condition associated with a person's mental well-being. Often referred to as the "ailment of modern times," projections suggest that by 2030, one in three individuals could grapple with depression disorder, commonly known as Major Depressive Disorder (MDD). According to the World Health Organization (WHO), approximately 280 million people worldwide are affected by this disorder. However, due to self-denial among certain patients and a lack of recognition in various regions, depression may persist undiagnosed or untreated, exacerbating the condition.

Various studies in the literature concur that social media platforms, where individuals freely share their thoughts and express emotions, can serve as a vital source for monitoring health issues and trends. Platforms like Facebook enable researchers to comprehend and explore various facets of psychological concerns and human behavior through user posts. Notably, in studies focused on mental depression, comments on posts related to Major Depressive Disorder have been identified as valuable indicators for predicting future depressive behavior in individuals.

A recent survey highlights a growing number of people exhibiting depression symptoms, particularly teenagers and young adults, turning to social media as an outlet to articulate their feelings. However, prevailing work in this domain often relies on specific keywords like 'depression' and 'diagnose' when utilizing data. This approach assumes that social media users experiencing depression would directly use such explicit terms, which may not be the case.

The strategies employed for depression detection have predominantly embraced a bottom-up approach, leveraging deep learning (DL) and machine learning (ML) methods. While advancements in Natural Language Processing (NLP) methods using DL have occurred, the predictive capacity of these approaches is constrained, mainly due to the limited scope of DL methods learning from extensive datasets. Additionally, hybrid methods, combining multiple approaches, and ensemble methods, where several learning methods are integrated, consistently demonstrate high performance in addressing this problem.

II. Related Works

A dataset sourced from Reddit in 2017 is publicly accessible, obtained through the Reddit Inc. API. It encompasses various attributes such as id, writing, title, and date. The dataset primarily comprises narratives submitted by individuals diagnosed with depression, serving as labeled data for depression. Additionally, a control group contributed their information, constituting non-depressed labeled data. The dataset involves 887 subjects, with 135 identified as experiencing depression. Similar to other survey-derived datasets, this collection combines data from a depression survey with publicly available information from social media.

This work reveals that use of language plays an important role in depression detection as words describe what someone's current mental state is.Most research investigations on the identification of depression are based on textual data or person-descriptive methods using social media posts to select features. The linguistic elements of the social media content, such as words, Part of Speech (POS), and other linguistic properties, are the subject of textual-based featuring. The descriptive-based featured technique focuses on subject descriptions, including age, gender, employment position, income, drug or alcohol consumption, smoking, and other personal information about the subject or patient. These features are then input into the detection models. Most models for depression detection have been developed using ML classifiers, such as the Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB), , Random Forest (RF), , Bagging Predictors (BP), and other single and ensemble models.Considering deep learning as the leading technology with the latest applications. Deep learning of successful activities that immediately identify text, image, or sound can be learned with computer models. Deep learning models achieved higher accuracies that sometimes exceeded human-level performance.

While our goal is a generalised approach for depression detection in social media texts, the datasets include only text messages and exclude any emoticons, emojis, pictures, videos and web links that are commonly part of social media messages.

A. Limitations of Existing Works

The review of existing literature extensively explored various methodologies in the realm of depression detection and prevention, with a focus on leveraging data mining techniques. Earlier frameworks primarily concentrated on identifying depression from limited phrases, sentences, and blog posts, resulting in less-than-optimal accuracy. Addressing these drawbacks, our framework integrates advanced data preprocessing and employs deep learning techniques for the early identification of depression, utilizing user-generated tweets. Our approach is designed to surpass the limitations of prior works and enhance overall accuracy. We utilize performance evaluation metrics for a thorough analysis, rigorously assessing and validating the efficacy of our proposed model. This contributes valuable insights to the domain of depression detection and prevention, specifically through the analysis of social media data.

III. Dataset

There is a publicly available dataset from Facebook, which we've had referred and used in our ML and DL models to train. , which included features like Text, Age, Age Category, label and Gender. This mostly included reports provided by people who were diagnosed with depression. This was considered as depression labelled data. The data consist of users who are suffering or might be diagnosed with depression it is having 7489 different users, out of which 1359 are suffering from depression and the rest are not.Fig. 1. will be showing about the users who are depressed or not.Similar to many other datasets obtained through surveys, this dataset combines information from a depression survey with data sourced from public social media platforms.

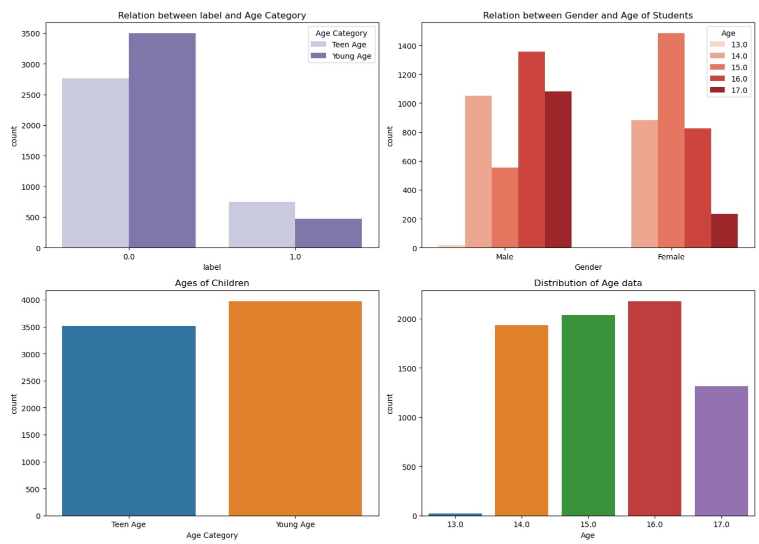

Fig. 1: Dataset

The collected dataset is having the age distributed from 13 - 17 which mainly consist of teen age and young age. Below Fig. 2. shows the distribution of user's within the specified range providing the visual representation count of users between the above mentioned bracket.

Dataset consist of Male and Female Gender which may be or may not be diagnosed with Depression. Below Fig. 3. represents the distribution of the age bracket which is 13-17 and out of which how many of the user's are Male and Female.

IV. Proposed Methodology

Given that the dataset used for training our models is imbalanced, with a significant amount of noise that could potentially impact the performance of our model, it becomes imperative to address these challenges. The dataset also includes unused or irrelevant data that may not contribute to accurately predicting the user's mental health status. In order to enhance the quality of our data and improve the precision of health predictions, we have implemented specific steps for data cleaning.

To mitigate the impact of imbalance and noise, our data cleaning process involves carefully identifying and filtering out irrelevant or redundant information. We employ robust techniques to balance the dataset, ensuring that the model is trained on a representative and unbiased sample. Additionally, we selectively eliminate data points that do not contribute meaningfully to the prediction of user mental health, enhancing the overall efficiency of our model.

By adhering to these data cleaning protocols, we aim to create a refined dataset that is conducive to accurate and reliable predictions, ultimately contributing to the credibility and effectiveness of our approach in evaluating users' mental health.

A. Data Preprocessing

This is first and foremost step while detecting depression on textual data. As the dataset which we referred is having so many liabilities which is not useful in prediction but it might effect the performance of ML and DL models. Whole data is converted to lowercase.

1) Removing html tags : HTML tags such as this `<.*?>` won't help us in predicting the mental health of user which is having symptoms of depression

2) Removing URL's : URL links which are used in comments or used while chatting are of no use in above text classification which is helping in predicting the user's health

3) Removing Punctuation Marks : Punctuation marks such as (.,:?!""-[,]) are nothing but useless in this context of the data

4) Expanding chat acronyms : In contemporary communication, users frequently employ short-form expressions, like "lol" for "Laugh out Loud" and "asap" for "As soon as possible," to streamline conversations on social media. Recognizing the potential for confusion in our predictive model due to these acronyms, we have implemented a meticulous preprocessing step. This involves systematically replacing such acronyms with their actual meanings. By doing so, we ensure that our model accurately interprets user messages, even when presented in abbreviated form. This approach enhances the model's ability to discern the nuances of modern communication, contributing to more precise predictions of users' mental health based on their social media interactions.

5) Removing Stop Words: While preprocessing the removal of stop words is also necessary.Some Stop words are (the,a,me,i) that occur frequently in conversation which carries little semantic meaning. By removing stop words, we can focus on the words that contribute more to overall meaning of a text.

6) Treatment of Emoji's : There are two approaches we can use in this case where we can remove the emojis or we can replace them with their original meaning.But replacing the emoji's with their actual meaning will be quiet helpful in training our models. When we demojize it gives the added advantage in correct predictions.

B. Tokenization

It is one of the crucial step while performing text classification and in the proposed case depression detection.Tokenization breaks down a text into individual words or tokens. This provides a more granular representation of the text, allowing the model to understand the meaning of the text at the word level. Each token becomes a feature that the model can use for classification.Tokenization standardizes the representation of text, making it easier to compare and analyze documents. It ensures that each piece of text is broken down into the same basic units, which contributes in consistent and reliable analysis.

C. Stemming

Here we've two apprach to convert words to their root form, one is Stemming and other is lemmatization. Depending on requirement we use either of these methods, here we are going with stemming due to it's performance for large datasets compare to lemmatization. Stemming is the process of reducing inflection in words to their root forms such as mapping a group of of words to the same stem even if the stem itself is not a valid word in the language. It is used when we need speed or want our job to be done quickly as possible.Advantage of doing this step before feature selection is the the words which are having same meaning are mapped to their root word and not treated as separate. If it is treated as different words even though context is same, it could lead to inaccurate representations and analysis.Here we are using Porter stemmer algorithm which was invented in 1980 by Martin F.Porter.

D. Feature Extraction

For Extraction of Features from text data set we are using TF-IDF vectorizer. TF-IDF(Term Frequency - Inverse Document Frequency)

After tokenizing TF is calculated which measures how often the tem is used in a specific document

IDF is calculated for each term in entire corpus.This measure helps to asses the importance of term in the context of the entire collection of the documents.Terms that are common across many documents receive lower IDF score,while terms that are more unique to specific documents receive higher IDF score.

- $\text{TF}(t,d) = \frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total number of terms in document } d}$
- $\text{IDF}(t,D) = \log\left(\frac{\text{Total number of documents in the corpus } N}{\text{Number of documents containing term } t}\right)$
- $\text{TF-IDF}(t,d,D) = \text{TF}(t,d) \times \text{IDF}(t,D)$

- t is the term (word) for which we are calculating the IDF.
- D is the collection of documents.
- N is the total number of documents in the corpus.
- $df(t,D)$ is the document frequency of the term t, i.e., the number of documents in which the term t occurs.

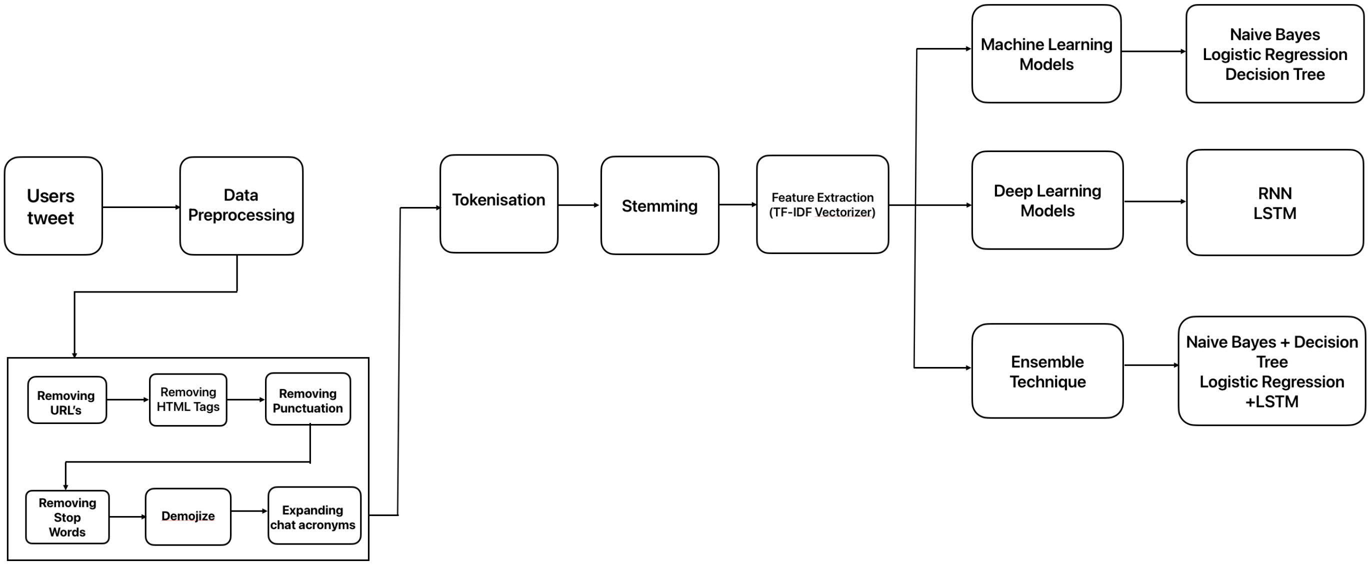

Fig. 2: Proposed Architecture

```
2762    So proud of dreamies ©ðŸ˜ they must be very happy
2763                           ywa ka bipolar ba nimo oyy
2764                              it's easy to make it viral
2765                 yuuhi no oyama ni, keep it up momiji
2766                                         Dizziness™ƒ
2767                                               gm tl
Name: text, dtype: object
```

Fig. 3: Before Preprocessing

```
2762    so proud of dreami they must be veri happi
2763                     ywa ka bipolar ba nimo oyi
2764                        it easi to make it viral
2765            yuuhi no oyama ni keep it up momiji
2766                                          dizzi
2767                          good morn talk later
Name: cleaned_tweet, dtype: object
```

Fig. 4: After Preprocessing

TF-IDF score is calculated by multiplying TF and IDF score. The process results in numerical representaion of importance of each term in document. Higher TF-IDF values indicates greater importance of a term in a specific document.

The TF-IDF scores for all terms form a feature vector for each document. These feature vectors can then be used as input for machine learning algorithms in text classification tasks such as Depression Detection. Each dimension in the vector corresponds to a unique term in the vocabulary, and the TF-IDF score represents the importance of that term in the document.

E. Training on various Machine Learning models

1) Naive Bayes Classifier: The Naive Bayes classifier is a supervised machine learning algorithm, which is used for classification tasks, like text classification. It is also part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category.

For each class, calculate the prior probability, which is the probability of a document belonging to that class without considering the features. $P(C_i)$ represents the prior probability of class C_i

For each term (feature) in the vocabulary and for each class, calculate the conditional probability of observing that term given the class label. This involves calculating $P(T_j|C_i)$, the probability of term T_j given class C_i.

2) Decision Tree: Decision Trees are a popular and intuitive machine learning algorithm used classification. They work by recursively partitioning the data into subsets based on the most significant attribute at each step. The result is a tree-like structure where each internal node represents a decision based on a particular feature.

The algorithm evaluates different features and selects the one that best splits the data into homogeneous subsets based on the target variable. The chosen feature is used to split the data into subsets at each internal node, and the process is repeated for each subset. This recursive partitioning continues until a stopping criterion is met. To make a prediction for a new data point, it traverses the tree from the root to a leaf node based on the feature values of the data point. Without proper control, Decision Trees are prone to overfitting, capturing noise in the training data.

3) Logistic Regression: Logistic Regression is a statistical method used for binary classification problems, where the outcome variable is categorical and has two classes (usually denoted as 0 and 1). Despite its name, logistic regression is a classification algorithm, not a regression algorithm. It estimates the probability that a given input belongs to a particular category. The logistic regression model is based on the logistic function (also known as the sigmoid function). The logistic function is defined as:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

where: - $\sigma(z)$ is the logistic (sigmoid) function. - e is the base of the natural logarithm. - z is a linear combination of the input features.

In logistic regression, the linear combination z is defined as:

$$z = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n$$

where: - b_0 is the intercept. - $b_1, b_2, \ldots, b_n$ are the coefficients associated with each feature $x_1, x_2, \ldots, x_n$.

The logistic regression equation can be written as:

$$P(Y=1) = \sigma(b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n)$$

Here, $P(Y = 1)$ represents the probability of the dependent variable (Y) being equal to 1.

F. Deep Learning Models

1) Recurrent Neural Network(RNN): Recurrent Neural Networks (RNNs) are a type of neural network architecture commonly used for sequential data, making them suitable for text classification tasks. In text classification, the goal is to assign predefined categories or labels to text documents. Learning rate used is 0.001 and batch size used is 4096. At input layer Activation Function used is Rectified Layer Unit(ReLu) to introduce the non linearity . At Output layer Sigmoid Activation Function is used.

2) Long Short Term Memory(LSTM): LSTMs are a type of recurrent neural network (RNN) designed to address the vanishing gradient problem in traditional RNNs. They are well-suited for handling sequences of data, making them effective for tasks like text classification.

LSTMs have a unique architecture that includes a memory cell. This memory cell allows the network to capture and remember information over long sequences, making it particularly useful for understanding the context of words in text.Learning rate used is 0.0001 and batch size used is 512. At input layer Activation Function used is Rectified Layer Unit(ReLu) to introduce the non linearity . At Output layer Sigmoid Activation Function is used.

G. Ensemble Techniques

Ensemble techniques in text classification involve combining the predictions of multiple base classifiers to improve overall performance. These methods aim to trigger the strengths of different models and mitigate their individual weaknesses. Here are some popular ensemble techniques used in text classification.

1) Voting Classifier: Combine the predictions of multiple base classifiers through voting (e.g., majority or weighted voting).

Hard Voting : Where the classification is based on majority number models giving same output.

Soft Voting : Where the classifiers provide probability estimates for each class of outcome

Naive Baye's with Decision Tree :
In these model we have used both Machine Learning models Naive Bayes and Decision Tree with depth 3. W've used Voting Classifier and used Hard voting technique for the implementation.

LSTM + Logistic Regression :
In order to implement 2 or more models together we've used ensemble technique in which we used 'SOFT' Voting Classifier.'SOFT' Voting Classifier is based on probability estimates for each outcomes. The implemented ensemble technique of LSTM + Logistic Regression outperforms the other implemented models.The results are shown in the below table.

H. Performance Measure

The performance measurement of datasets is carried out with the help of a confusion matrix from where we get the results, which calculates the values of accuracy, precision, and recall with the help of positive and negative values of datasets. The formulas used for the calculation of values are,

True Positive (TP): Number of correctly predicted positive instances.

True Negative (TN): Number of correctly predicted negative instances.

False Positive (FP): Number of incorrectly predicted positive instances (Type I error).

False Negative (FN): Number of incorrectly predicted negative instances (Type II error).

1) Accuracy: The ratio of correctly predicted instances to the total instances.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

2) Precision: The ratio of correctly predicted positive instances to the total predicted positive instances.

$$\text{Precision} = \frac{TP}{TP+FP}$$

3) Recall: The ratio of correctly predicted positive instances to the total actual positive instances.

$$\text{Recall} = \frac{TP}{TP+FN}$$

4) $F1_Score$: The harmonic mean of precision and recall, providing a balance between the two.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

V. Experimental Results

The results suggest that deep learning models, specifically LSTM, outperform traditional machine learning models such as Naive Bayes and Decision Tree in terms of accuracy, precision, F1 score, and recall. Logistic Regression also demonstrates high performance, indicating its effectiveness in text classification tasks.

The standout performance of the Long Short-Term Memory (LSTM) network, as evidenced by its high accuracy (94.92%), precision (93.97%), F1 score (94.71%), and recall (94.73%), suggests that this deep learning model excels in capturing the nuanced language indicative of depressive tendencies within social media comments. The ability of LSTM to analyze and understand the context of comments enables a more accurate and sensitive detection of individuals at risk of depression.

In comparison to traditional machine learning models like Naive Bayes and Decision Tree, which show reasonable performance, the superiority of LSTM becomes more pronounced in the context of social media comments. Logistic Regression, with its accuracy of 93.45% and strong precision, F1 score, and recall, also proves to be effective in identifying potential cases of depression based on comments.

The hybrid model, combining LSTM and Logistic Regression (LSTM+LR), achieves a remarkable accuracy of 97.35%. This suggests that the fusion of deep learning capabilities with traditional machine learning approaches can lead to heightened accuracy in binary classification tasks related to depression detection in social media comments.

These findings underscore the potential of advanced computational models in scrutinizing social media comments for signs of depression, thereby providing a valuable tool for early identification and intervention. As social media continues to be a significant platform for self-expression, these results contribute to the ongoing efforts in utilizing technology to support mental health monitoring and care.

Classifiers	Accuracy (%)	Precision (%)	F1 Score (%)	Recall (%)
Naive Baye's	67.35	81.77	71.42	67.35
Logistic Regression	93.45	93.72	92.85	93.45
Naive Baye's + DT	88.45	89.85	85.75	88.45
DT	87.51	89.14	84.19	87.51
RNN	92.92	92.06	91.02	91.92
LSTM	93.92	93.97	94.71	94.73
LSTM+LR	96.35	96.35	95.27	93.34

TABLE I: Result

VI. Conclusion

Depression is one of the most common mental disorders permeating worldwide. It is important to educate ourselves about depression on an individual, communal, and global scale. Addressing the issue and helping individuals suffering from depression should be given utmost priority. Our proposed models are Naive Baye's Decision Tree,Logistic Regression,RNN,LSTM,Naive Baye's + DT and LSTM + Logistic Regression. Out of these LSTM and LSTM + Logistic Regression out performs other with accuracy of 94.92% and 97.10%.

As mentioned in the report we've used different techniques to detect the depression on text data. There are other techniques which can be used in text classification BERT technique can also be used or we can use different combination of models which can be fused with Ensemble Technique.

References

[1] B. H. Hidaka, "Depression as a disease of modernity: Explanations for increasing prevalence" J. Affect. Disorders, vol. 140, no. 3, pp. 205–214, Nov. 2012

[2] S. Ji, C. P. Yu, S.-F. Fung, S. Pan, and G. Long,"Supervised learning for suicidal ideation detection in online user content," Complexity, vol. 2018, pp. 1–10, Sep. 2018.

[3] Luna Ansari , Shaoxiong Ji , Qian Chen, and Erik Cambria,"Ensemble Hybrid Learning Methods for Automated Depression Detection",on Computational Social System, VOL. 10, NO. 1, FEBRUARY 2023.

[4] Nafiz Al Asad ,Md. Appel Mahmud Pranto,Sadia Afreen,Md. Maynul Islam,"Depression Detection by Analyzing Social Media Posts of User", Communication Systems(SPICSCON) 28-30 November,2019.

[5] G. K. Gupta and D. K. Sharma, "Depression Detection on Social Media with the Aid of Machine Learning Platform: A Comprehensive Survey," 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2021, pp. 658-662.

[6] B. Zohuri and S. Zadeh, "The Utility of Artificial Intelligence for Mood Analysis, Depression Detection, and Suicide Risk Management," Journal of Health Science, vol. 8, pp.67-73, 2020.

[7] L. Ma. and Y. Wang, "Constructing a Semantic Graph with Depression Symptoms Extraction from Twitter", in Proc. of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2019, pp.1-5.

[8] X. Tao, X. Zhou, J. Zhang, and J. Yong, "Sentiment analysis for depression detection on social networks," in Proc. of the International Conference on Advanced Data Mining and Applications, 2016, pp. 807-810.

[9] A. H. Orabi, P. Buddhitha, M. H. Orabi, and I. Diana, "Deep learning for depression detection of Twitter users," in Proc. of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, June 2018, pp. 88-97.

[10] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in Reddit social media forum," IEEE Access, vol. 7, pp. 44883-44893, 2019